Roadmap Epigenomic Tutorial
WashU EpiGenome Browser

Keystone Symposia - Epigenomics & DNA methylation

March 30, 2015 – Keystone, Colorado.

Presenter: Vasavi Sundaram

Tutorial Overview:

1. Getting started with the EpiGenome Browser.......................................2
2. Loading data on the EpiGenome Browser...3
3. Navigating the EpiGenome Browser..4
4. Sessions: saving and retrieving browsing sessions...............................12
5. Using Apps on the EpiGenome Browser..14

Notes:

- To follow along with this tutorial, use instructions marked by ➤ . We have provided screenshots for guidance (ordered by bracketed alphabets: [A], [B], …)
- Click in the order of the numbered-circles on the screenshots to follow this tutorial.

- To catch up with any section of the tutorial, use the provided session ID and the session name, listed at the end of that section. For this, follow these steps:
    - ➤ [A] Click on the **Apps** menu, and then click on the **Sessions** banner.
    - ➤ [B] Under the **Retrieve** button, enter the session ID and then click on the **Retrieve** button.
    - ➤ This will generate a list of session names, from which click on the session of interest.

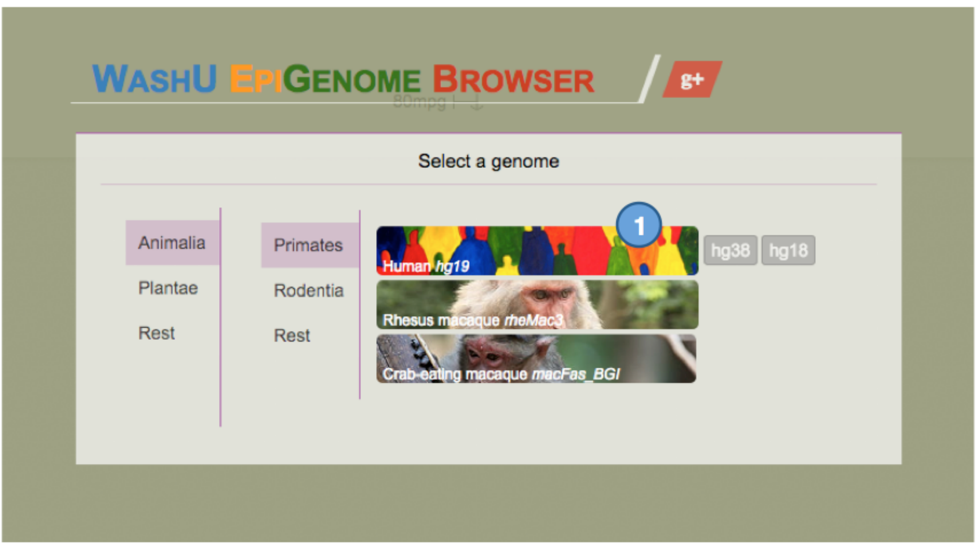
1. Getting started with the EpiGenome Browser

1.1. Load the browser

- Access the browser at http://epigenomegateway.wustl.edu/browser

1.2. Select the genome assembly of interest

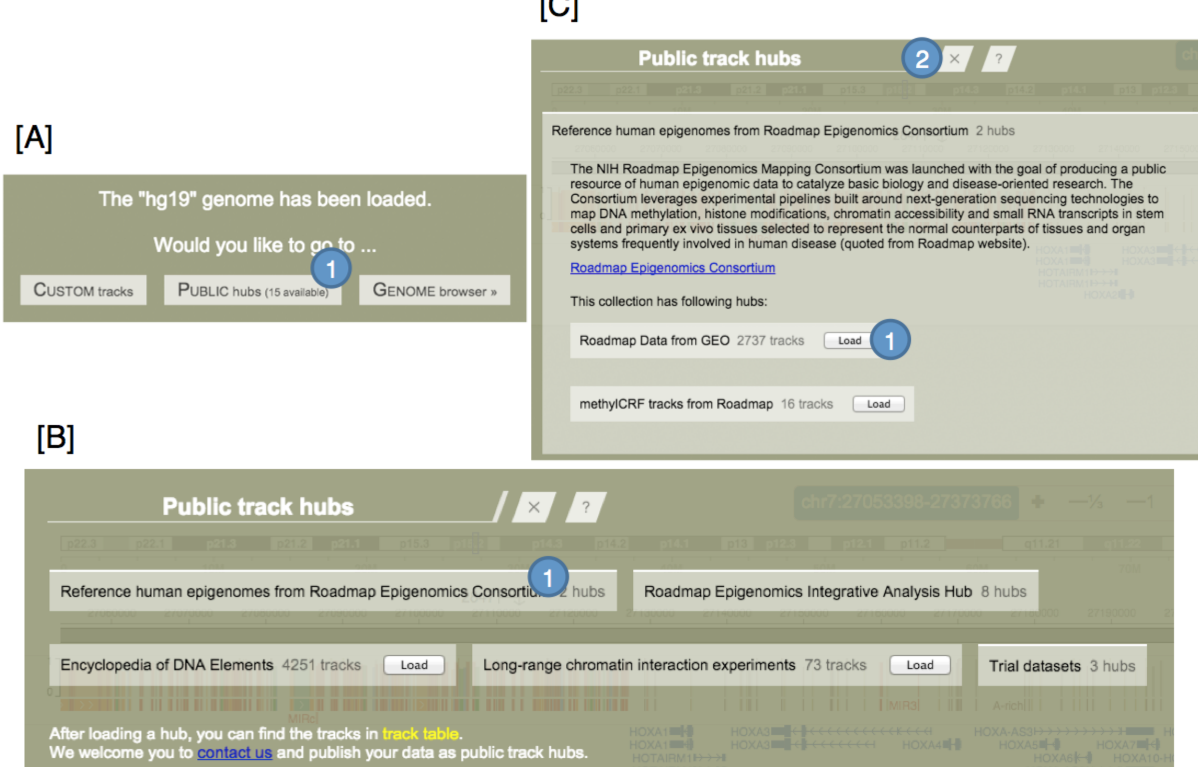
- For the purpose of this tutorial, please select Human hg19.



2. Loading data on the EpiGenome Browser

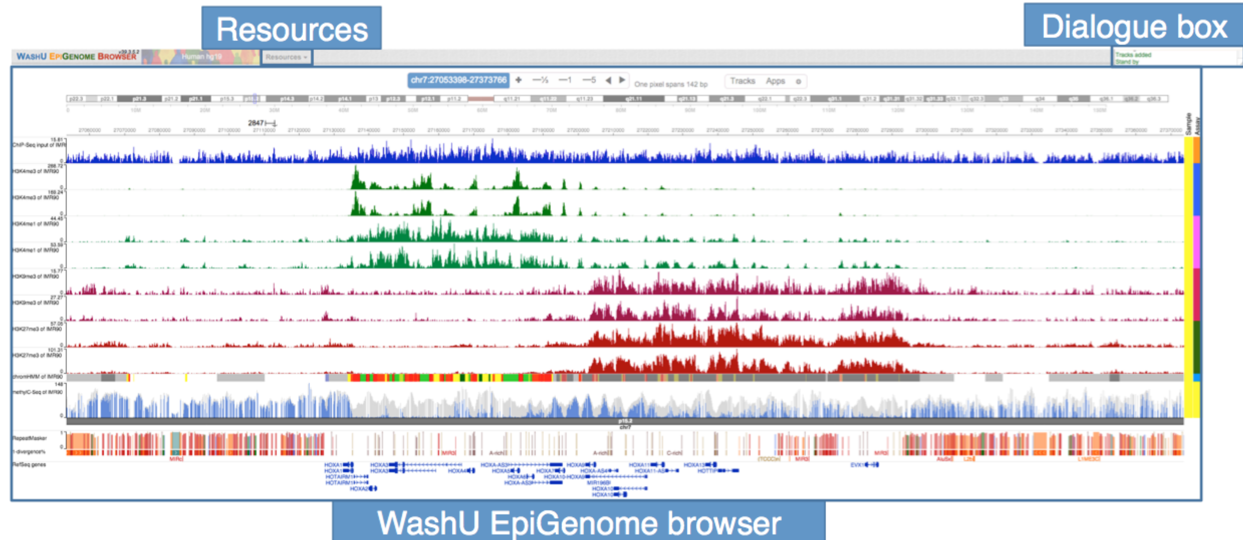
2.1. Data hubs: A data hub is a collection of tracks/datasets that can be viewed on the browser.

- [A] Click on the `PUBLIC hubs (15 available)` banner to view all the available public datasets on the EpiGenome Browser.
- [B] Click on the `Reference human epigenomes from Roadmap Epigenomics Consortium` banner.
- [C] This will generate a list of available hubs. Click the `Load` button on `Roadmap Data from GEO` box. Once the datasets are loaded, `exit` the data hub section by clicking on the `X` at the top-right of the floating window, or pressing “`Esc`”.



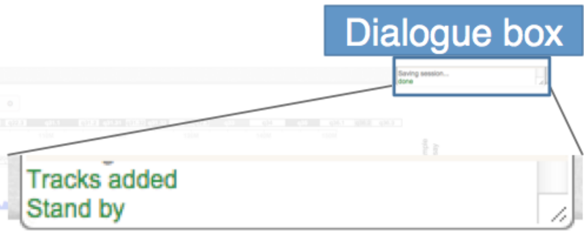
3. Navigating the EpiGenome Browser

3.1. Layout of the EpiGenome Browser



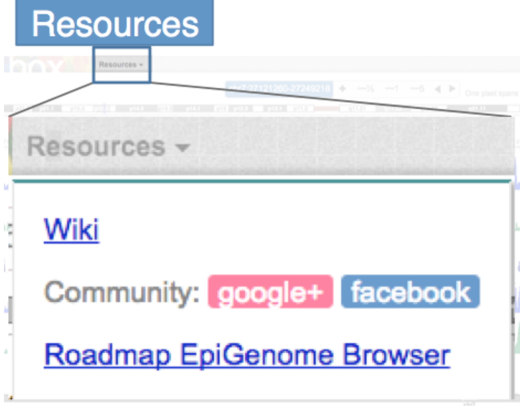
3.1.1. Dialogue box

➢ This box logs all the actions performed by the EpiGenome browser. It is useful for knowing of any error encountered by the browser while performing a task.



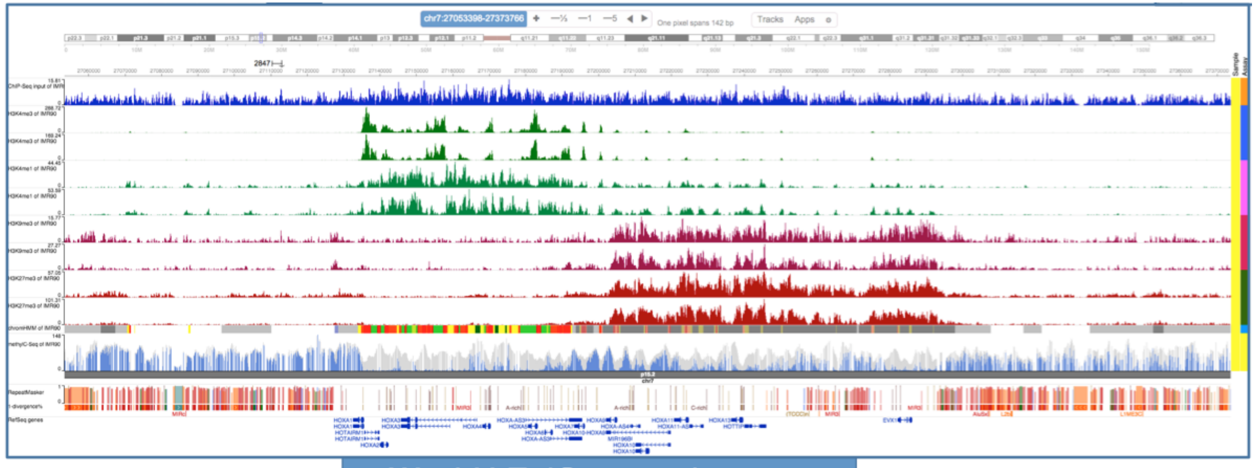
3.1.2. Resources

➢ This window contains links to: (i) the documentation on the browser (Wiki), (ii) social-media discussion forums (on Google+ and Facebook) for the browser, and (iii) the Roadmap EpiGenome Browser



3.1.3. EpiGenome Browser

- This region contains the data tracks aligned to the genome, overlaid on gene annotations and other annotations customized by the user.

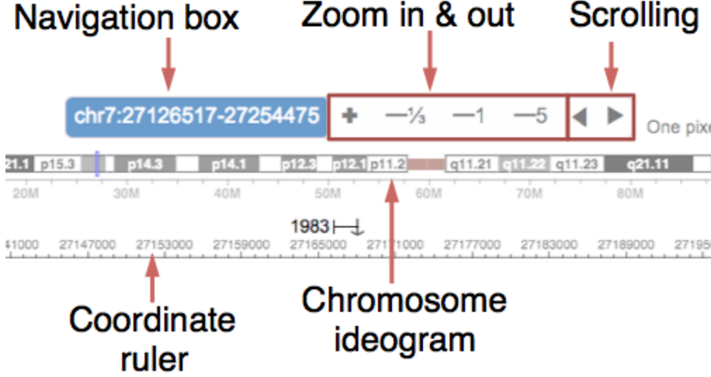


WashU EpiGenome browser

3.2. Navigating the genome on the EpiGenome Browser

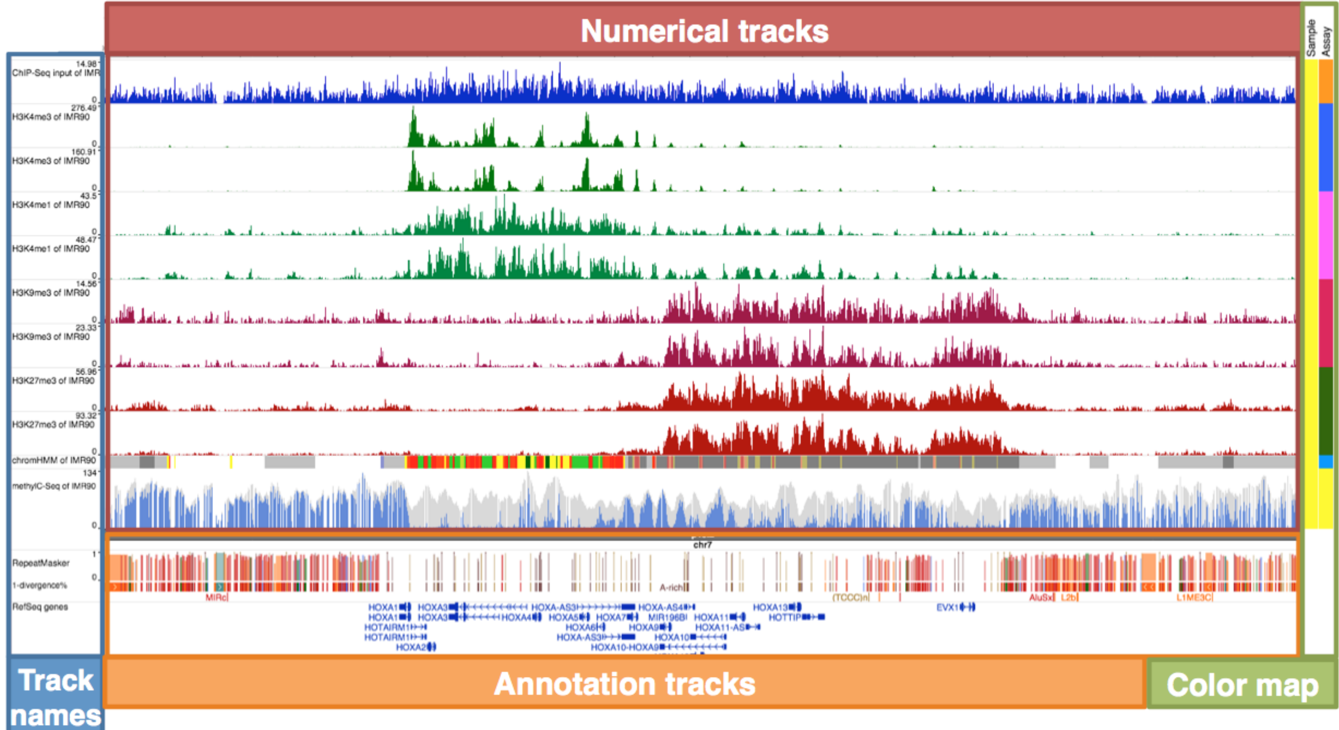
3.2.1. Genome navigation controls

- The controls enable moving across the genome, using the **zoom-in**, **zoom-out**, and **scrolling** buttons.
- Alternatively, click on the **navigation box** to enter the genomic coordinates of the region of interest. Click on **Go** to move to the new region.



3.2.2. Data visualization

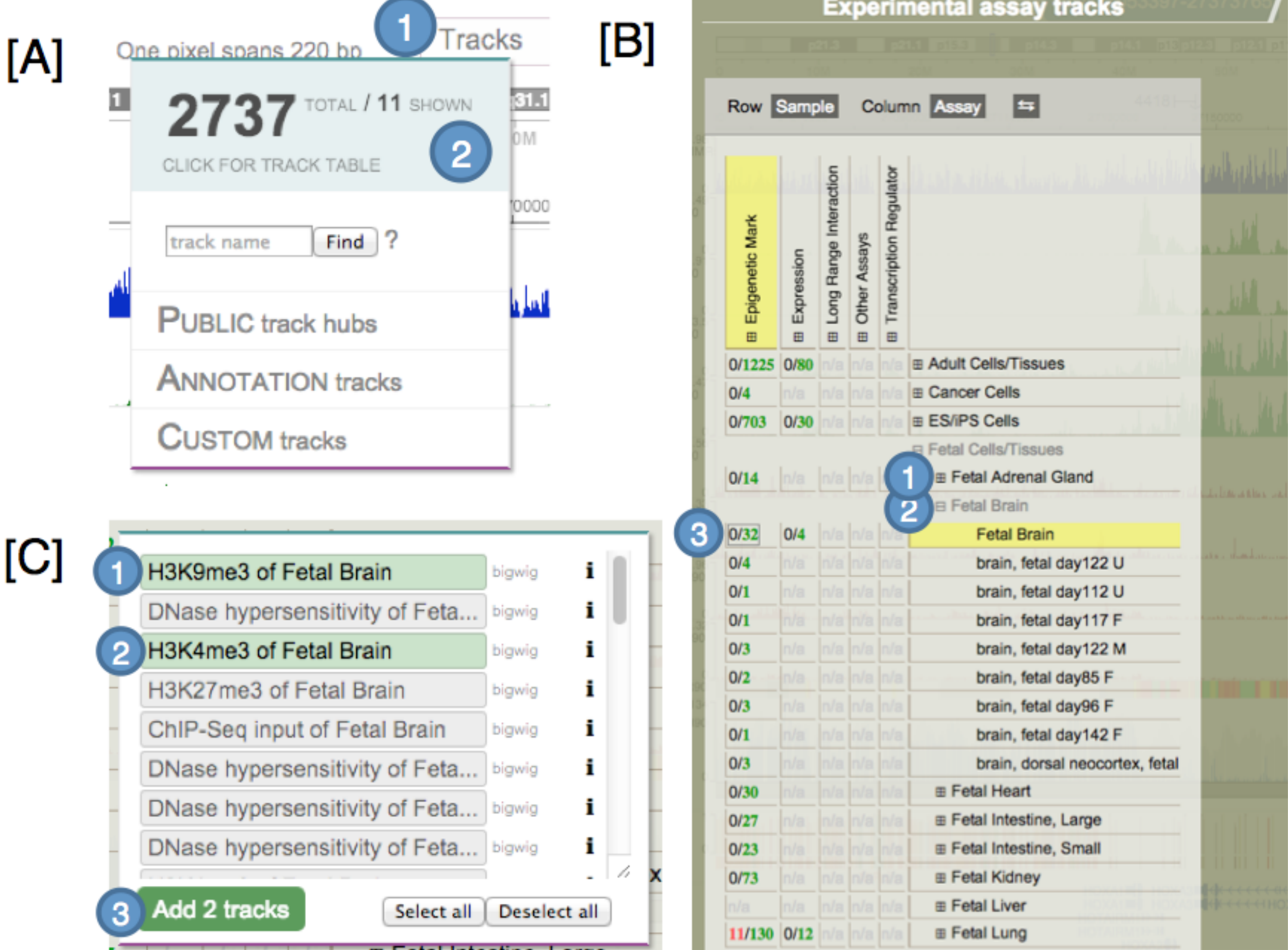
- **Numerical tracks** represent the density of reads from sequencing experiments aligned to the genome assembly.
- **Track names** for each track are listed on the left-hand side of the track itself.
- On the right-hand side of the tracks is a **color map**, which represents the **metadata** for the tracks.
- Below the numerical tracks are **annotation tracks**, including annotations of genes and transposable elements.



3.3. Searching for datasets on the EpiGenome Browser

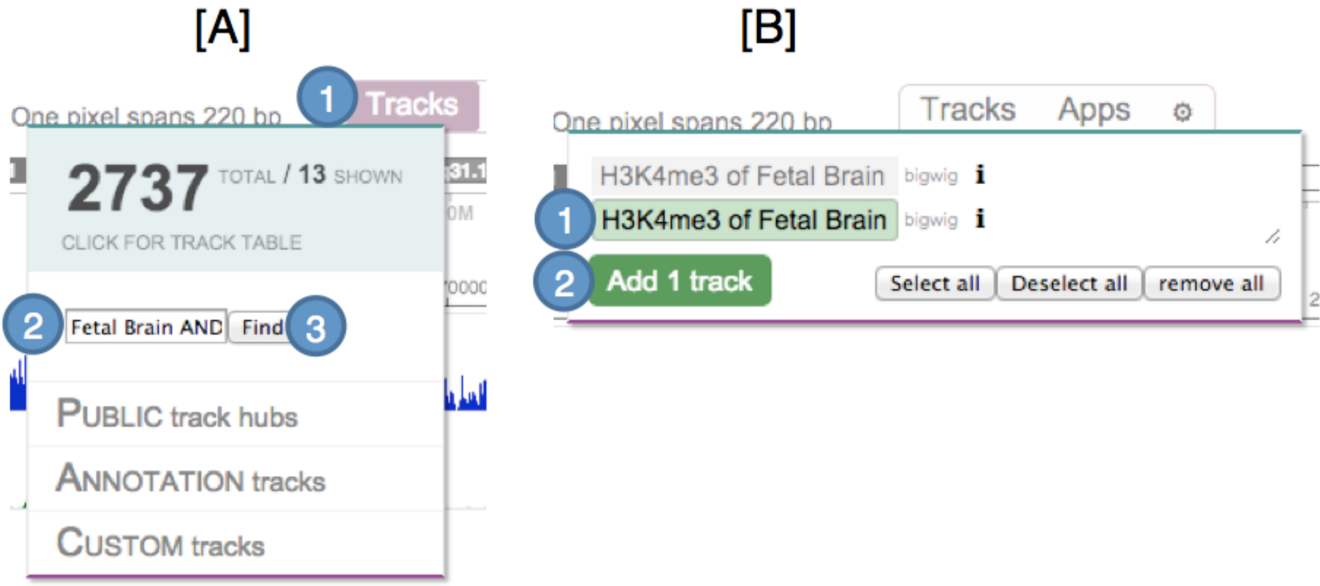
3.3.1. Facet table: The facet table organizes all the loaded datasets into a table, to allow the user to search for datasets using metadata terms. Here, rows represent samples and columns represent assay types. Cells with numbers represent the number of datasets. For example, the cell with numbers: 11/130 - here, 11 represents the number of datasets currently loaded on the browser, while 130 represents the total number of datasets available.

- [A] To display the facet table click on the Tracks menu, and then click on the teal-highlighted box (at the top of the floating window).
- [B] This will generate the facet table. To add more datasets, click on Fetal Cells/Tissues and then select Fetal Brain. Click on the cell corresponding to the Epigenetic Mark column and the Fetal Brain row to list the datasets in this category.
- [C] To select data, first select the datasets H3K9me3 of Fetal Brain and H3K4me3 of Fetal Brain. Then, click on the button labeled Add 2 tracks.
- To exit this floating window, click on the X at the top-right of the window, or press Esc.



3.3.2. Searching for data using the track search box

- [A] Click on the Tracks menu. In the search box, type Fetal Brain AND H3K4me3, and then click on Find.
- [B] This will list all replicates of the dataset. Only one of the two boxes listed can be clicked, since the other replicate has already been loaded (previous section 3.3.1). Click on the second H3K4me3 and Fetal Brain, and then click Add 1 track.
- Repeat this process to add one more replicate track for Fetal Brain AND H3K9me3.



➢ Click outside the floating window to get back to the browser.

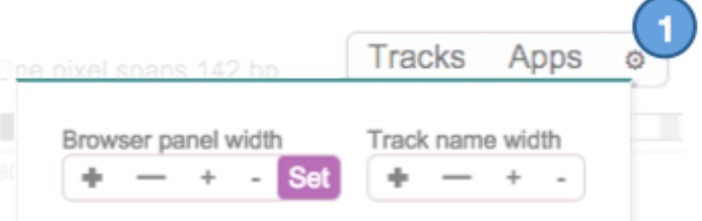
❖ To catch up with this point of the tutorial, enter the session ID - REMCdemo and click on the session named - REMC - IMR90 and Fetal Brain

3.4. Configuring the EpiGenome Browser

3.4.1. Formatting the view range

➢ The browser width can be adjusted using the + and - buttons under settings option (gear button). Click Set to change the browser width.

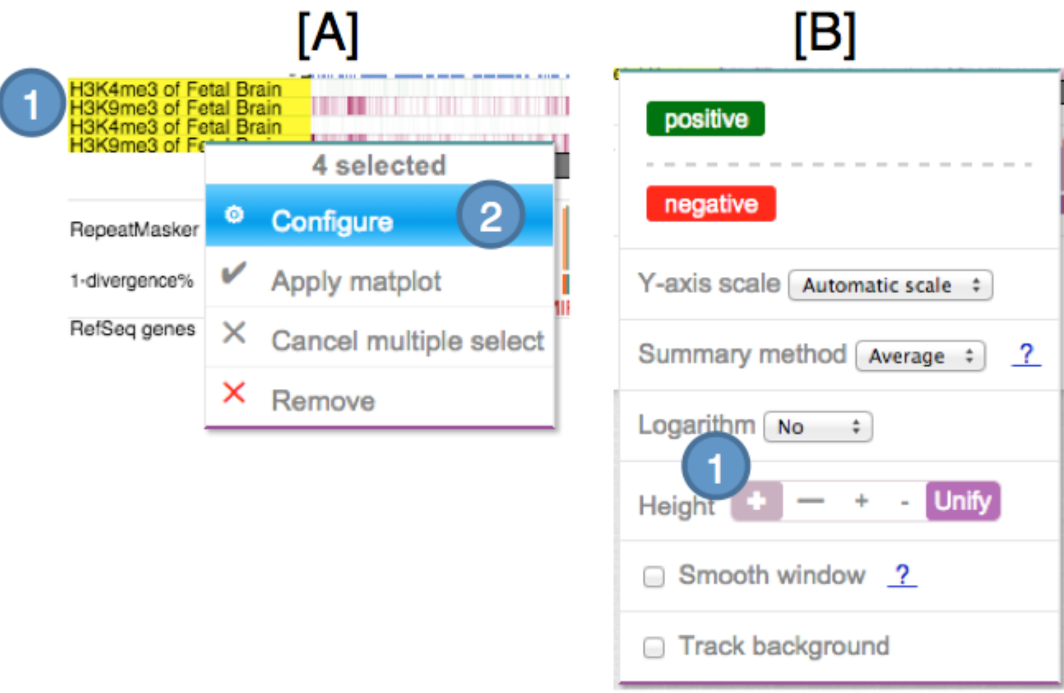
➢ Similarly, the width of the track names can be changed using the + and - buttons under the settings option.



3.4.2. Changing the height of tracks

➢ [A] To select multiple tracks, hold down the shift-key and click on the names of the 4 Fetal Brain tracks that was just added. Right-click on the yellow-highlighted track names. This will create a floating window; click the Configure menu.

➢ [B] Click on the + sign repeatedly to increase the height of the tracks to your preference.



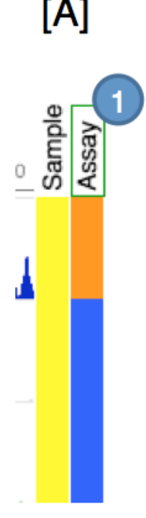
- ➢ Click outside this floating box to get back to the browser.
- ➢ To de-select the 4 tracks, right-click on the yellow-highlighted track names and select `Cancel multiple select`.

❖ To catch up with this point of the tutorial, enter the session ID - `REMCdemo` and click on the session named – Changed height of Fetal Brain tracks

3.4.3. Re-ordering tracks

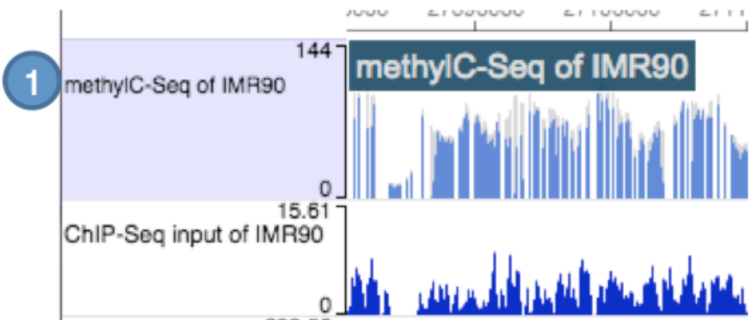
3.4.3.1. Reordering by metadata terms

- ➢ [A] To reorder the tracks based on assay type, click on `Assay` above the color map.



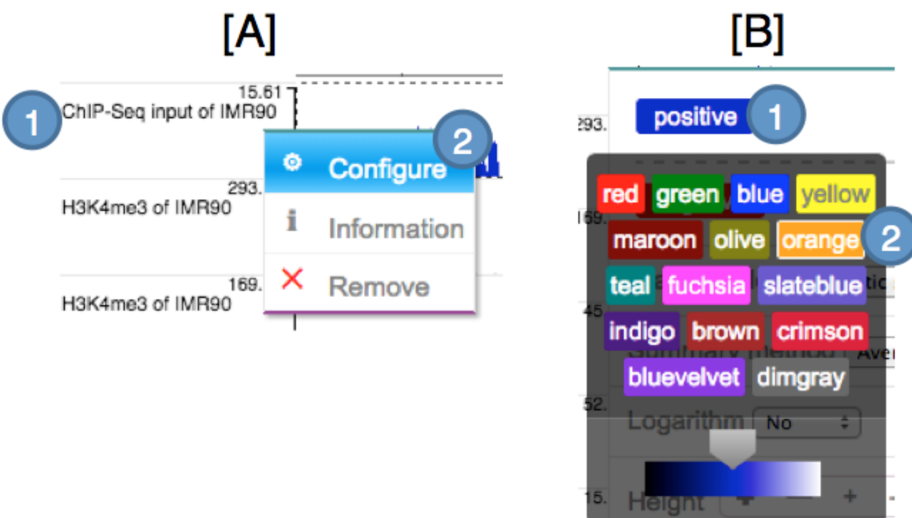
3.4.3.2. Reordering by click-and-drag

- Alternatively, clicking on a track name and moving that track to a new position can also reorder tracks.
- Select the methylC-Seq of IMR90 track and move the track above the chromHMM track.



3.4.4. Changing the color of tracks

- [A] Select the Chip-Seq input of IMR90 tracks by right-clicking on the track name and then click the Configure button.
- [B] In the Configure menu, click on the positive button and then select the color of your choice to change the color of the track.



- Click outside this floating box to get back to the browser.

❖ To catch up with this point of the tutorial, enter the session ID - REMCdemo and click on the session named – Changed color of ChIP-seq Input track

3.4.5. Changing the y-axis scale for tracks

- [A] Right-click on the blue color-map (under Assay metadata term) and click on Configure, to select the four H3K4me3 tracks. Click on Configure to change the height of the tracks.

➢ [B] Click on the **Y-axis scale** drop-down menu and then click **Fixed**. This will generate text boxes to enter the y-axis range. Enter **150** in the max text-box. Click on **apply** to change the y-axis scale.

➢ [C] Click on the button labeled **beyond threshold**. This will generate a floating window with color options; change the color to any color of your choice.

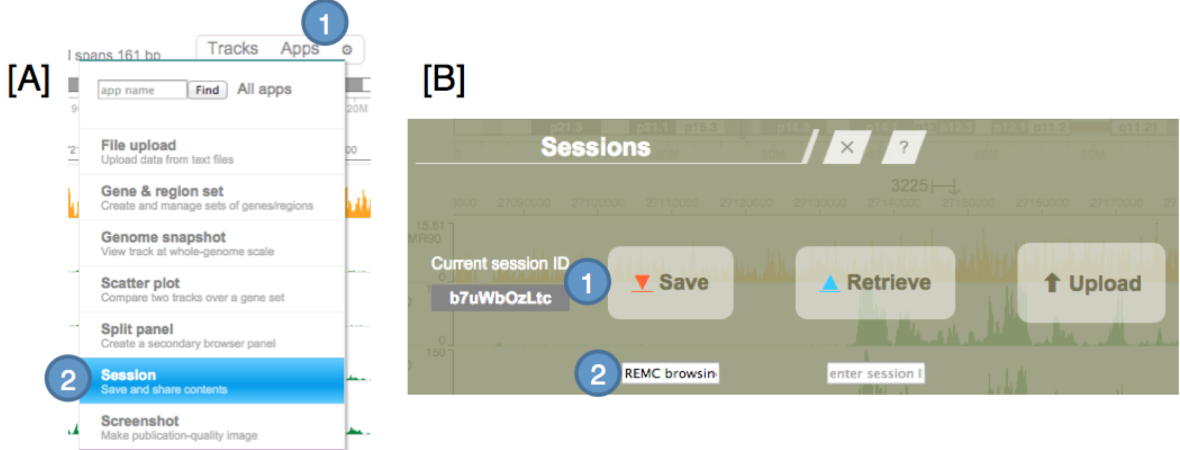
➢ Click anywhere outside the **Configure** box to get back to the browser.

❖ To catch up with this point of the tutorial, enter the session ID - REMCdemo and click on the session named – Set y-scale of H3K4me3 tracks

4. Sessions: saving and retrieving browsing sessions

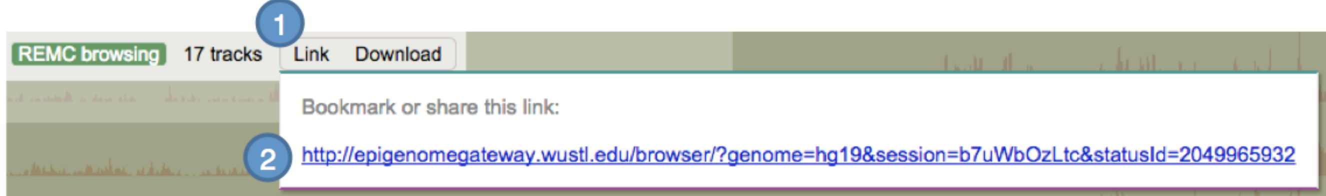
4.1. Saving sessions

- [A] Click on the Apps menu, and then select Session.
- [B] Enter "REMC browsing" in the dialogue box under the Save button, and then click Save.



4.2. Sharing links for collaboration

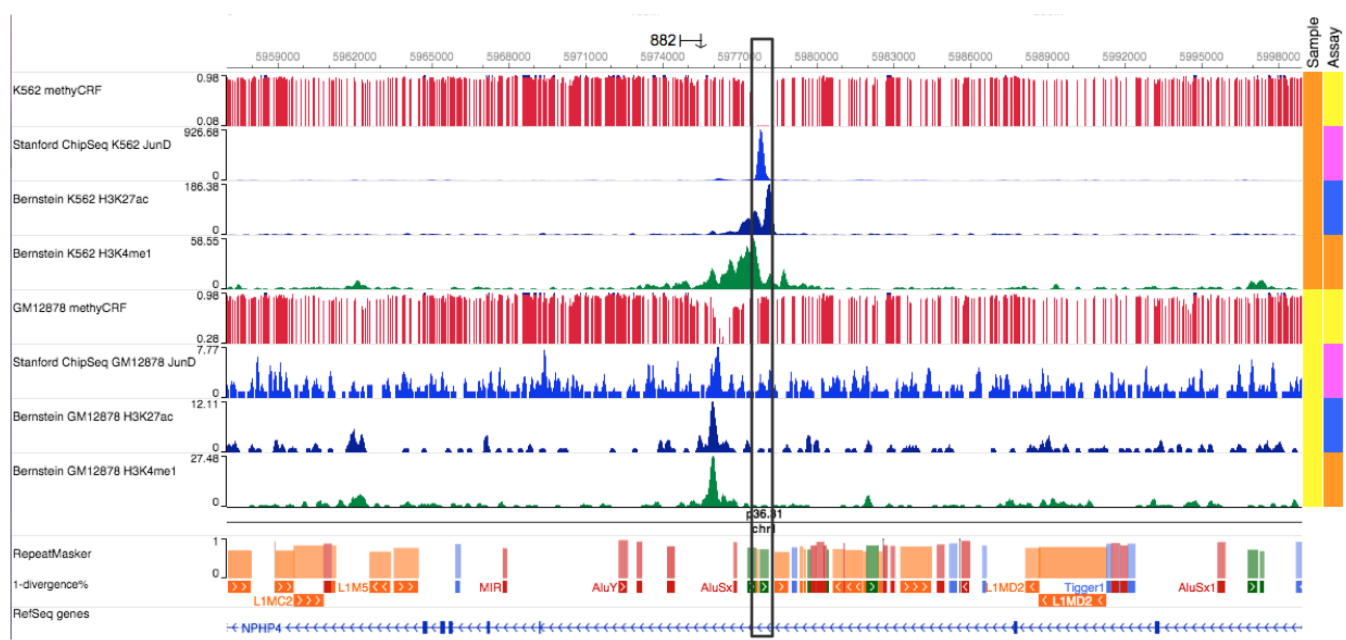
- To share the link of this session, click on the Link button beside the session name. This will create a floating window with a URL that can be shared



4.3. Retrieving sessions

- Launch a new instance of the EpiGenome browser in a new window (URL: http://epigenomegateway.wustl.edu/browser). Select the Human hg19 genome assembly. Click on Genome Browser >> to proceed to the browser.
- [A] Click on the Apps menu, and then select Sessions.
- [B] Enter the ID ENCODEdemo in the search box under the Retrieve button, and then click Retrieve. This will list all the available session under this ID. Click on MLT1
- Click on the X at the top-right of the floating window, or press Esc to get back to the browser.

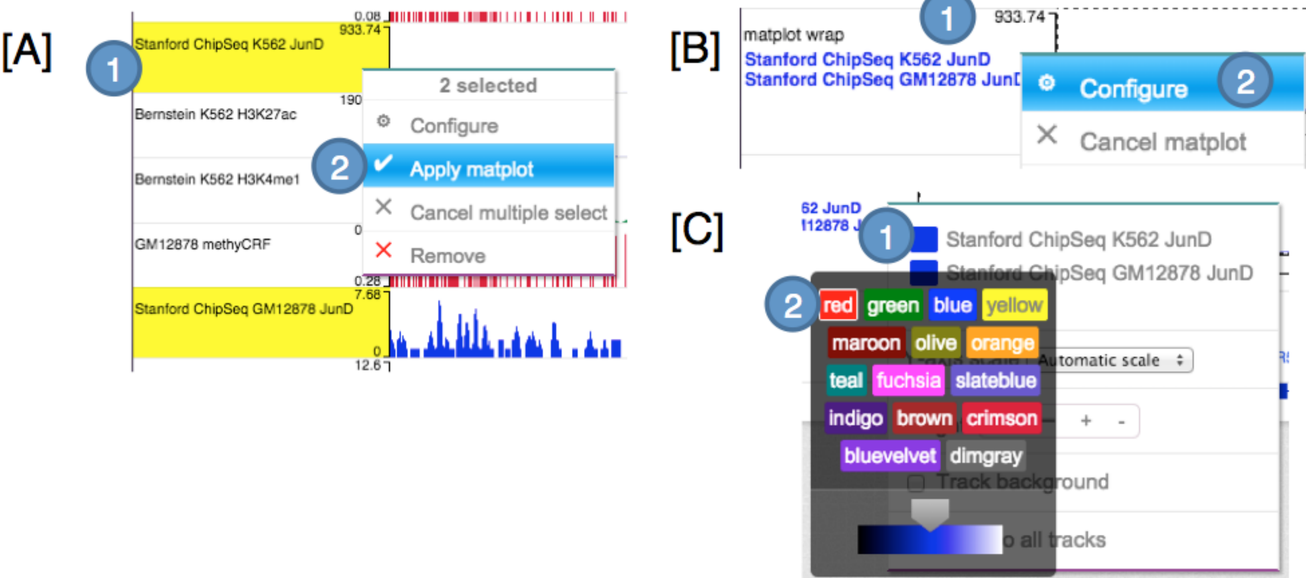
- This will retrieve a new session that will be used for the last section of this tutorial. Select **MLT1** to launch a new session.
- In this session, we will be using ENCODE datasets to analyze the cell-type specificity of JunD binding on transposable elements.
- The **MLT1** session shows the specific binding of JunD to a transposable element, MLT1 in K562, specifically. This is corroborated by K562-specific hypomethylation of MLT1, and GM12878-specific hypermethylation.



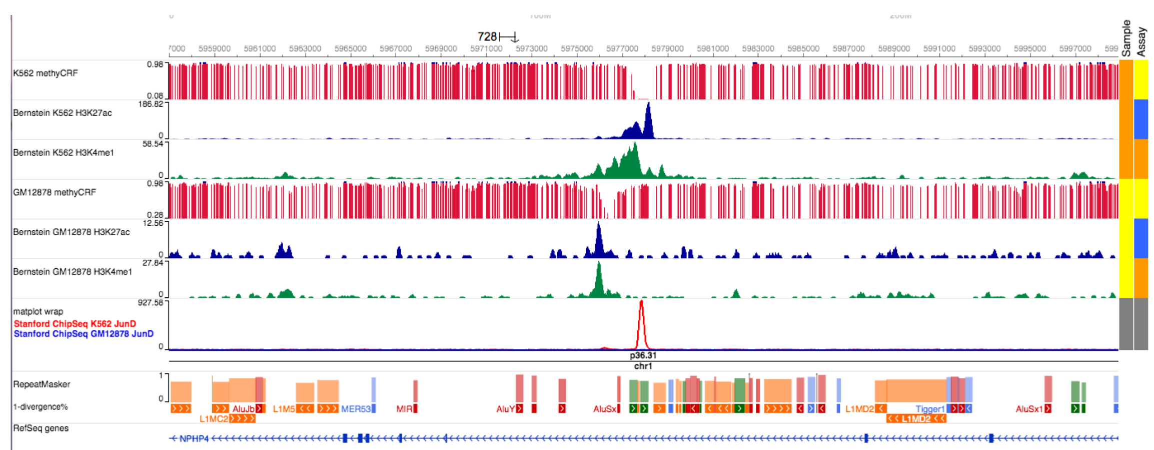
5. Using Apps on the EpiGenome Browser

5.1. Matplot: Compare two or more data tracks by plotting datasets to one y-axis scale.

- [A] Select the two `JunD ChipSeq` tracks and right-click on any of the tracks' names. Click on `Apply matplot`.
- [B] To change the color of the numerical tracks, right-click on the Matplot's track name, and click on `Configure`.
- [C] Click on the blue-box beside the K562 JunD label. Select `red` to change the color of the K562-JunD binding track from blue to red.



- This results in a track where the ChipSeq data for JunD binding in K562 and GM12878 share the same y-axis, and can be easily compared. It is evident from this view, that there is much more JunD binding in K562 compared with GM12878, on this MLT1 element.

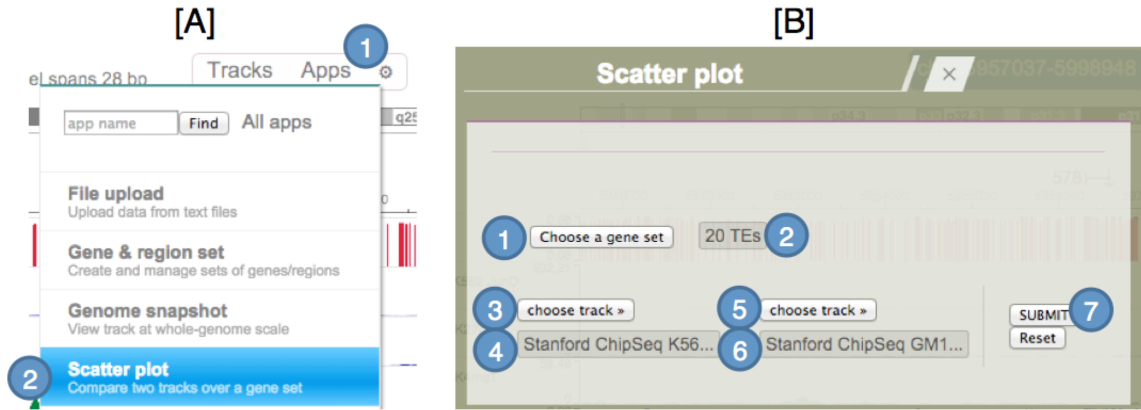


- ➢ To exit from the matplot view, right-click on the matplot track name and click on `Cancel matplot`.

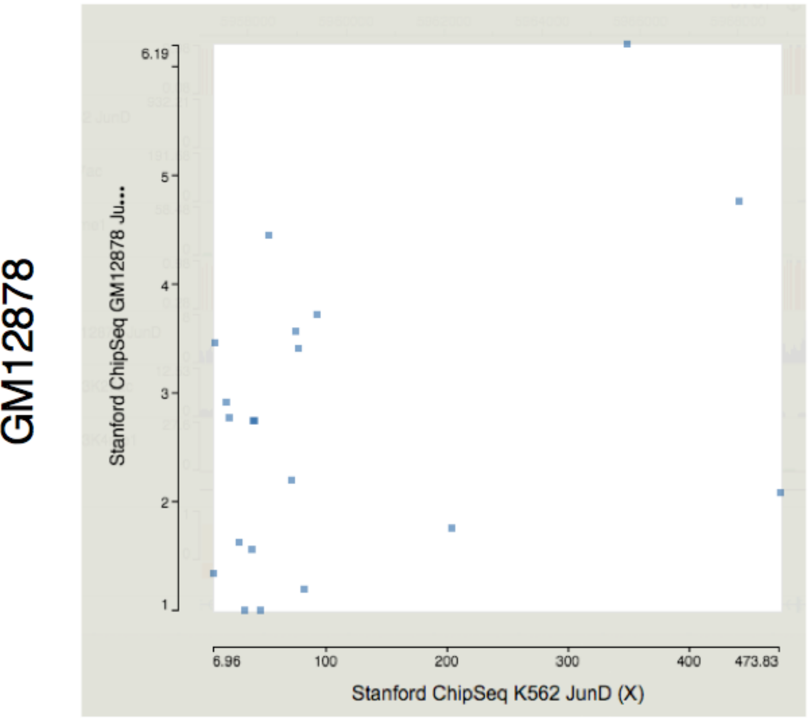
❖ To catch up with this point of the tutorial, enter the session ID - ENCODEdemo and click on the session named – MLT1-matplot

5.2. Scatter plot: Correlate different datasets, across multiple genomic regions.

- ➢ [A] Click on the `Apps` menu and then select `Scatter plot`.
- ➢ [B] To select the genomic regions to plot, click on `Choose a gene set` button and then click on the gene-set titled `20 TEs`. To select numerical tracks to plot, first click on the `Choose track` button and then select the dataset.
- ➢ [B] First select `Stanford ChipSeq K562 JunD` track and then select the `Stanford ChipSeq GM12878 JunD` track. Click on `Submit` to create the plot.



- The resulting scatter plot shows the correlation of ChipSeq signal for JunD binding in K562 (y-axis) and GM12878 (x-axis), wherein each transposable element (TE) is represented by a dot. Looking closer at the scatter plot it is evident that for most data points on the plot, the K562 ChipSeq signal is higher than the GM12878 ChipSeq signal, for JunD.



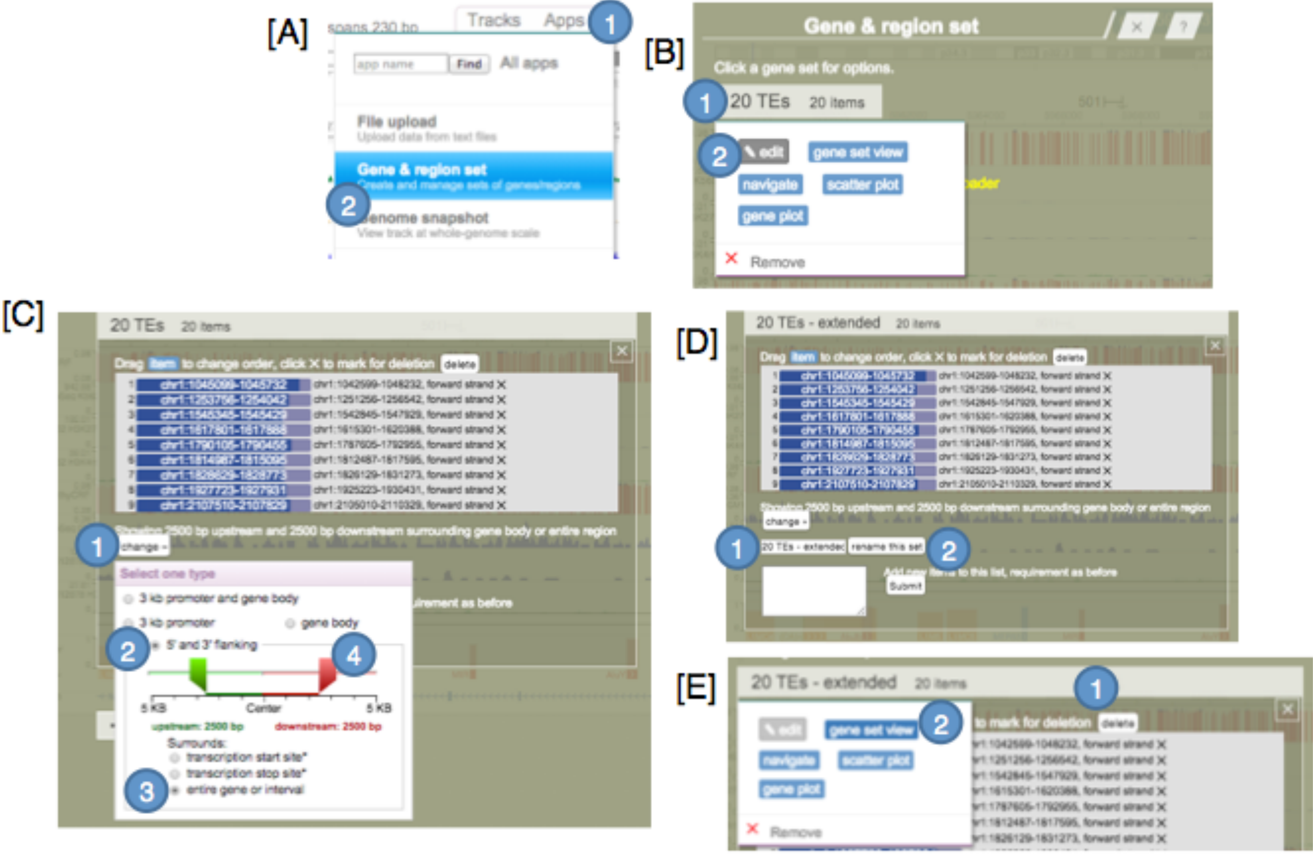
K562

- ➢ Click on **Go back** button to go back and create another scatter plot.
- ➢ Click on **X** or press **Esc** to exit the scatter plot app.

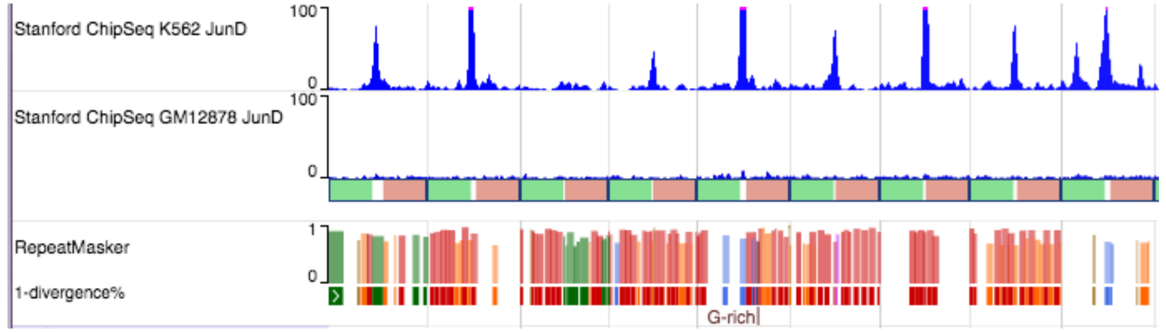
❖ To catch up with this point of the tutorial, enter the session ID - ENCODEdemo and click on the session named – Gene set loaded. This contains the pre-loaded session with the 20 transposable elements (named **20 TEs**) used in this demo. Follow the instructions under **Scatter plot** to create this plot.

5.3. Gene-set: Visualizing multiple genomic regions in parallel.

- ➢ [A] To view data on multiple genomic loci at a time, click on the **Apps** menu and then select **Gene & region set**.
- ➢ [B] Select the banner labeled "**20 TEs**" and click on **edit**.
- ➢ [C] Under the list of TEs, click on the **change** button to modify the view range. First select the radio button for **5' and 3' flanking** and then select the radio button for **entire gene or interval**. Adjust the red downstream cursor to 2500bp. Click anywhere outside the window to exit this window.
- ➢ [D] To rename this set, enter "**20 TEs – ext 2500bp**" and click on **Rename this set**.
- ➢ [E] To view this gene-set, click on the banner of the gene set and then click "**gene set view**".



- This results in a tiled-view of all the TEs arranged beside each other. By setting the same y-axis scale for the two JunD ChipSeq tracks, the K562-specificity of JunD binding on these TEs is evident.



➢ To exit the gene-set view click on the Genome navigation box, which is now labeled **Showing entire set | X**. This will lead back to the EpiGenome Browser.

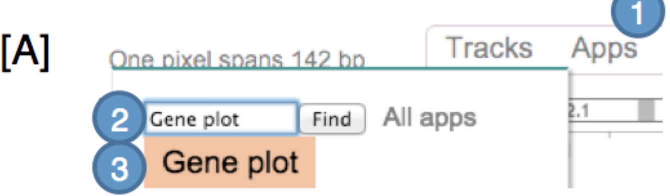
❖ To re-do this section of the tutorial, enter the session ID - ENCODEdemo and click on the session named – Gene set loaded. This contains the pre-loaded session with the 20 transposable elements used in this demo. Follow the instructions under **Gene-set** to create this plot.

❖ To view the gene-set on the EpiGenome Browser, enter the session ID - ENCODEdemo and click on the session named – Gene-set view. This will lead to the gene-set view.

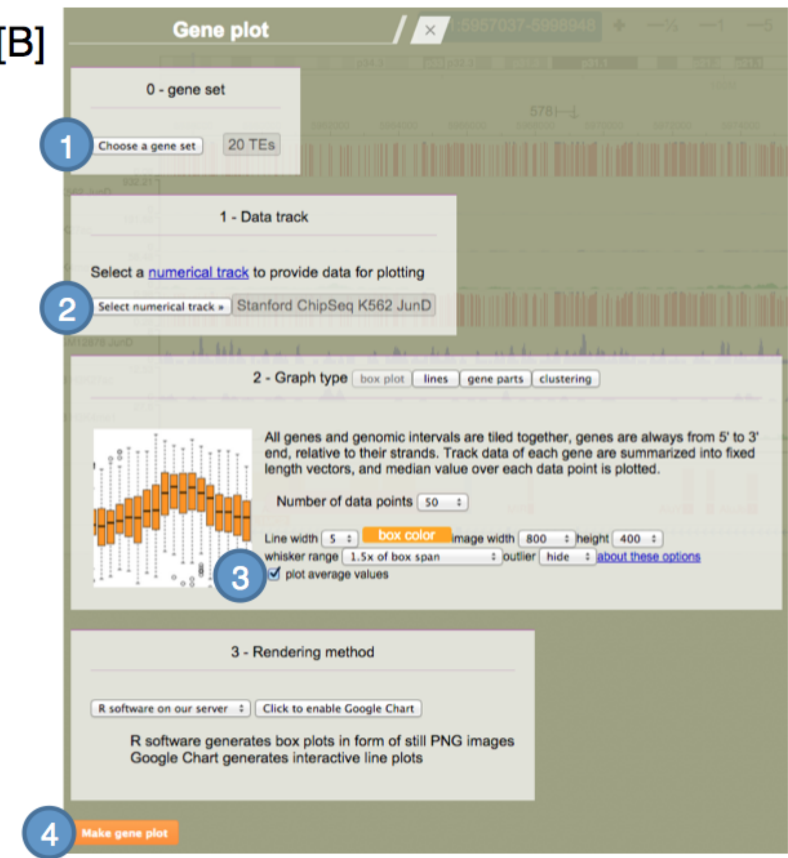
5.4. Gene plot: Summarize the data distribution across multiple regions, by splitting the regions into bins.

- [A] Click on the `Apps` menu. In the search box, type "`Gene plot`" and select the `Gene plot` app.
- [B] To make a gene plot, follow these steps:
    - In the `0 – Gene set` section - click on the `Choose a gene set` button. Select the `20 TEs` list of TEs for this analysis.
    - In the `1 – Data track` section – click on the `Select numerical track >>` button, and then select `Stanford ChipSeq K562 JunD`.
    - In the `2 – Graph type` section – the `box plot` graph type is selected by default. Let's use this for now. Click the check-box beside `plot average values`.
    - Click on the `Make gene plot` button to create the gene plot.
    - The same process can be repeated for the `Stanford ChipSeq GM12878 JunD`.

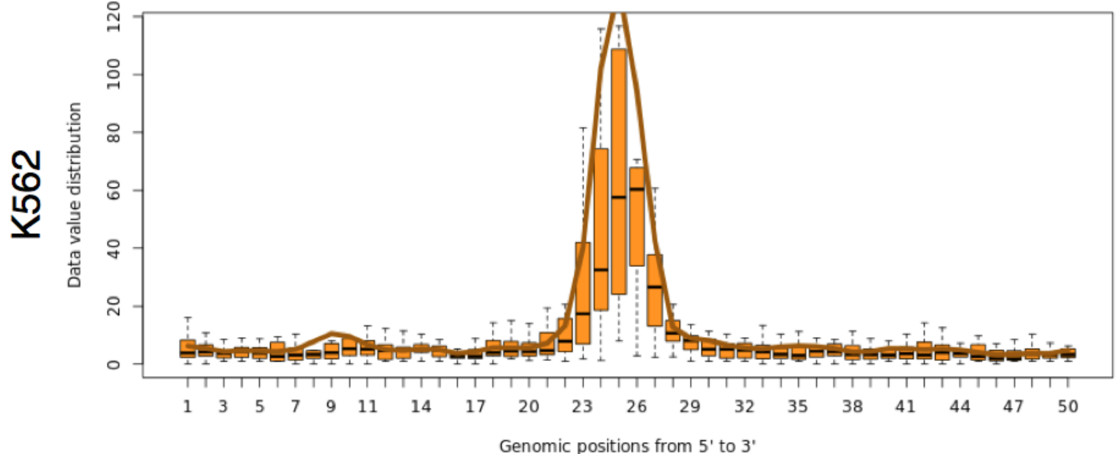
[A]

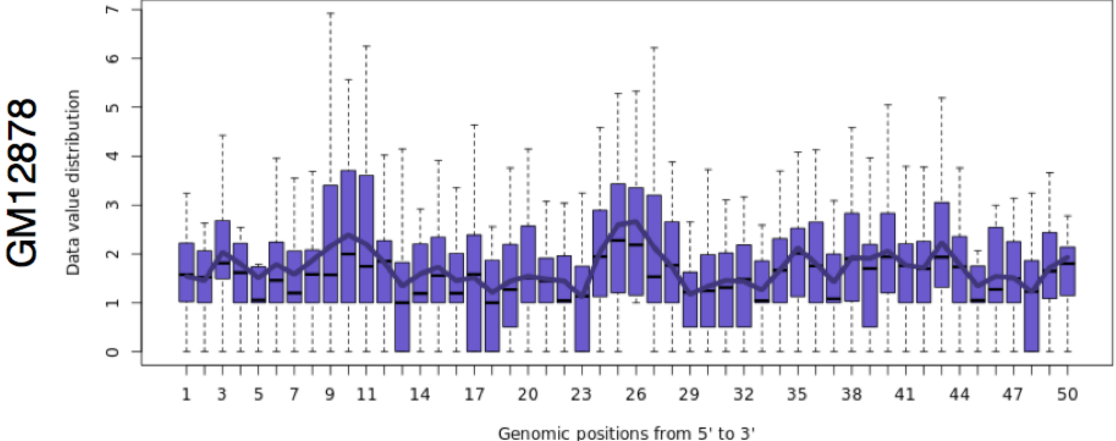


[B]



- Comparing the two gene-plots reveals that the highest ChIP-seq signal for JunD binding in K562 is on the TEs (data points 23-29 on the x-axis). The ChIP-seq signal for JunD binding in GM12878 on TEs is comparable to the flanking region (data points 1-23 and 29-50).





❖ To repeat this section of the tutorial, enter the session ID - ENCODEdemo and click on the session named – Gene set loaded. Follow the instructions under **Gene plot** to create this plot.

More information:

- For more documentation and tutorials, visit: http://epigenomegateway.wustl.edu/support/index.html

- Follow the WashU EpiGenome Browser on